AI and Machine Learning for Detecting Myocardial Infarction

Xiantong Xiang

Shandong University

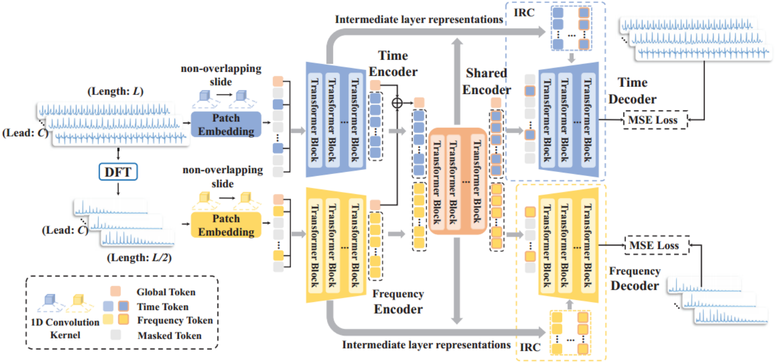
Tuesday, November 5th, 2024

Contents

Background

- High misdiagnosis rates for early diagnosis of cardiovascular disease and detection of potential patients
- Deep learning and the application of artificial intelligence
- Data and privacy issues
- Popularity of portable and wearable devices
- Electrocardiogram (ECG) is the most common, low-cost and convenient tool for diagnosing cardiovascular disease

- **Background**: the majority of methods concentrate solely on time domain information, overlooking the information originating from additional modalities or perspectives.
- **Method**: a novel bimodal masked autoencoder framework (BMIRC)
- **Innovation Point**:
  - a novel bimodal masked autoencoder framework for time-frequency joint modeling
  - internal representation connections (IRC) from the encoder to the decoder



- $ECG(Time) \stackrel{DFT}{\longrightarrow} ECG(Frequency)$
- $ECG(T\&F) \Rightarrow Encoder(T\&F) \Rightarrow Encoder(Shared) \Rightarrow Decoder \Leftarrow IRC$

- The tokens of T and F can be expressed as

$$Z_t = [z_t^1, z_t^2, \ldots, z_t^{\frac{L}{S}}] \in \mathbb{R}^{\frac{L}{S} \times D}$$

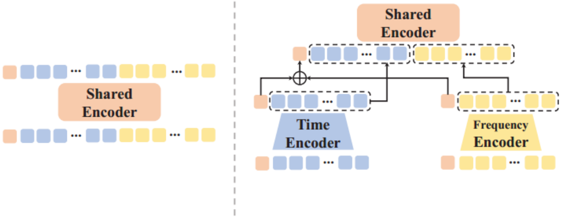
$$Z_f = [z_f^1, z_f^2, \ldots, z_f^{\frac{L}{2S}}] \in \mathbb{R}^{\frac{L}{2S} \times D}$$

- learnable position embeddings $\mathrm{PE} \in \mathbb{R}^{N \times D}$ are integrated into the patch embeddings

$$\tilde{I}_m = Z_m + \mathrm{PE}_m$$

$$I_m = \mathrm{Concat}(z_g^m, \tilde{I}_m)$$

- a random masking strategy, meaning that each token has the same probability of being masked.



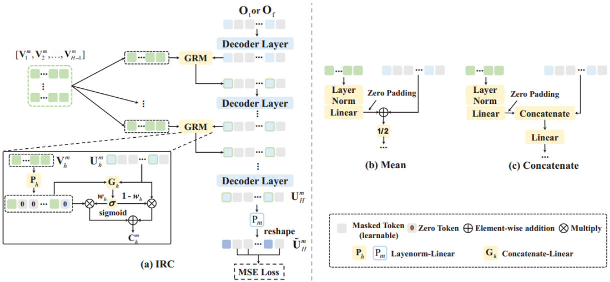
(a) All Shared (b) Specific-Shared(Ours)

- the global tokens of time and frequency modalities are added and inserted at the first position of the sequence, with the other tokens concatenated sequentially.

$$\tilde{O}_m = LN(O_m) = [\tilde{o}_g^m, \tilde{o}_1^m, \tilde{o}_2^m, \dots, \tilde{o}_n^m]$$

$$O_0^s = [o_g^t + \tilde{o}_g^f, \tilde{o}_1^t, \tilde{o}_2^t, \dots, \tilde{o}_n^t, \tilde{o}_1^f, \tilde{o}_2^f, \dots, \tilde{o}_n^f]$$

$$O_s = \Theta(O_0)$$



- gated representation mixer called GRM

$$\hat{V}_h = P_h(V_h)$$

$$w_h = \sigma(G_h(\hat{V}_h, U_h^m))$$

$$C_h^m = w_h * \hat{V}_h + (1 - w_h) * U_h^m$$

$$U_{h+1}^m = \Lambda_{h+1}(C_h^m)$$

Discussion

- Multimodal Data
- Missing
- Data Leakage
- DWT
- Matrices or Images